ATTACHMENT F

Exhibit E

Mitglied der Helmholtz-Gemeinschaft

JÜLICH
FORSCHUNGSZENTRUM

Project Sketch

DECO

Dynamical Exa-Computing

Exascale Computing Meeting, Brussels, 2.9.2010

Thomas Lippert for the DECO Consortium

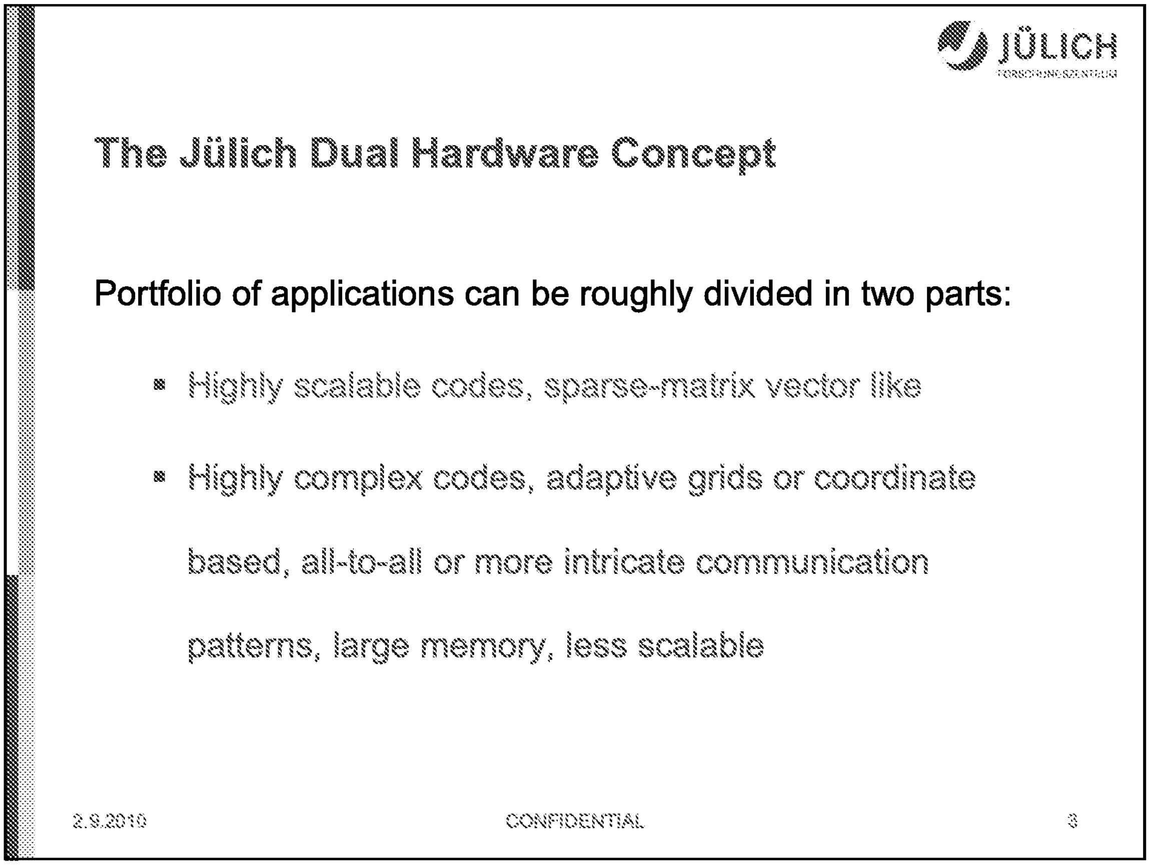
Exhibit E

JÜLICH
FORSCHUNGSZENTRUM

APPLICATION-INSPIRED APPROACH TOWARDS THE EXASCALE

2.9.2010 CONFIDENTIAL 2

Exhibit E



JÜLICH
FORSCHUNGSZENTRUM

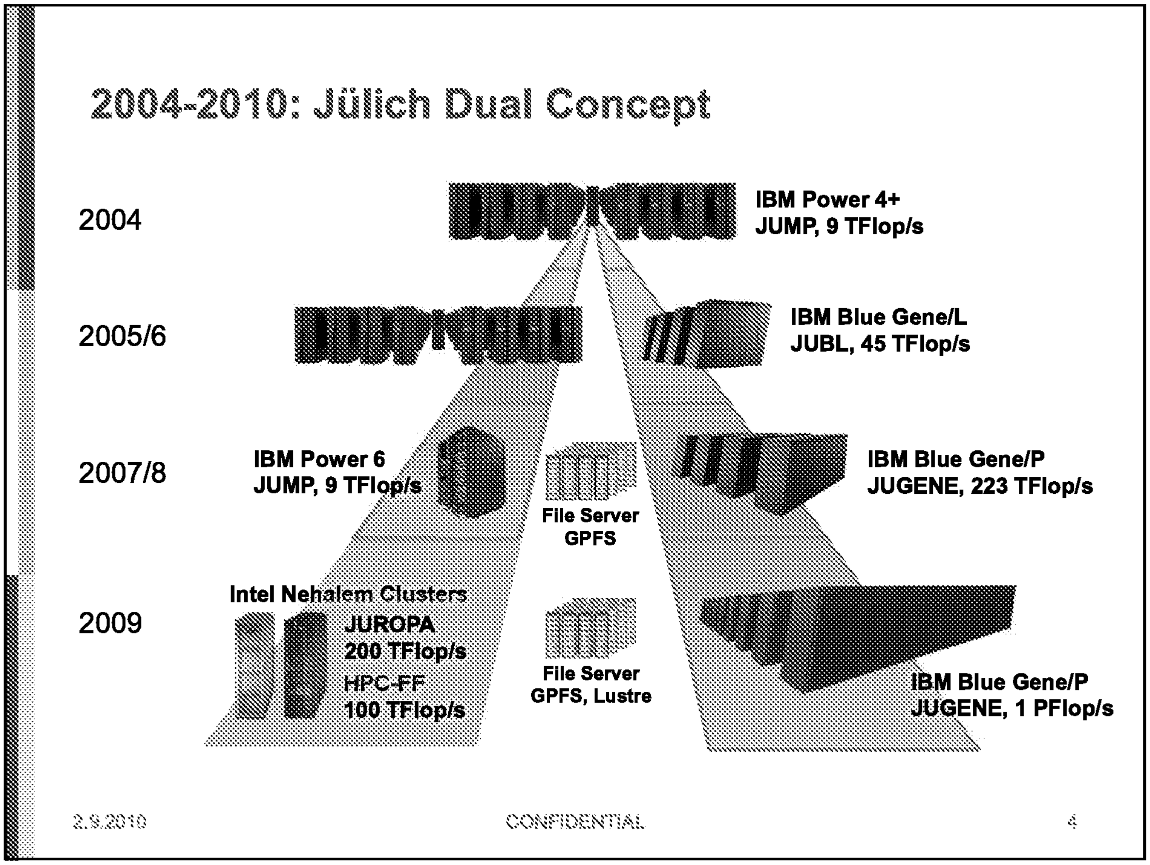
The Jülich Dual Hardware Concept

Portfolio of applications can be roughly divided in two parts:

- Highly scalable codes, sparse-matrix vector like
- Highly complex codes, adaptive grids or coordinate based, all-to-all or more intricate communication patterns, large memory, less scalable

2.9.2010 CONFIDENTIAL 3

Exhibit E



2004-2010: Jülich Dual Concept

2004

IBM Power 4+
JUMP, 9 TFlop/s

2005/6

IBM Blue Gene/L
JUBL, 45 TFlop/s

2007/8

IBM Power 6
JUMP, 9 TFlop/s

File Server
GPFS

IBM Blue Gene/P
JUGENE, 223 TFlop/s

2009

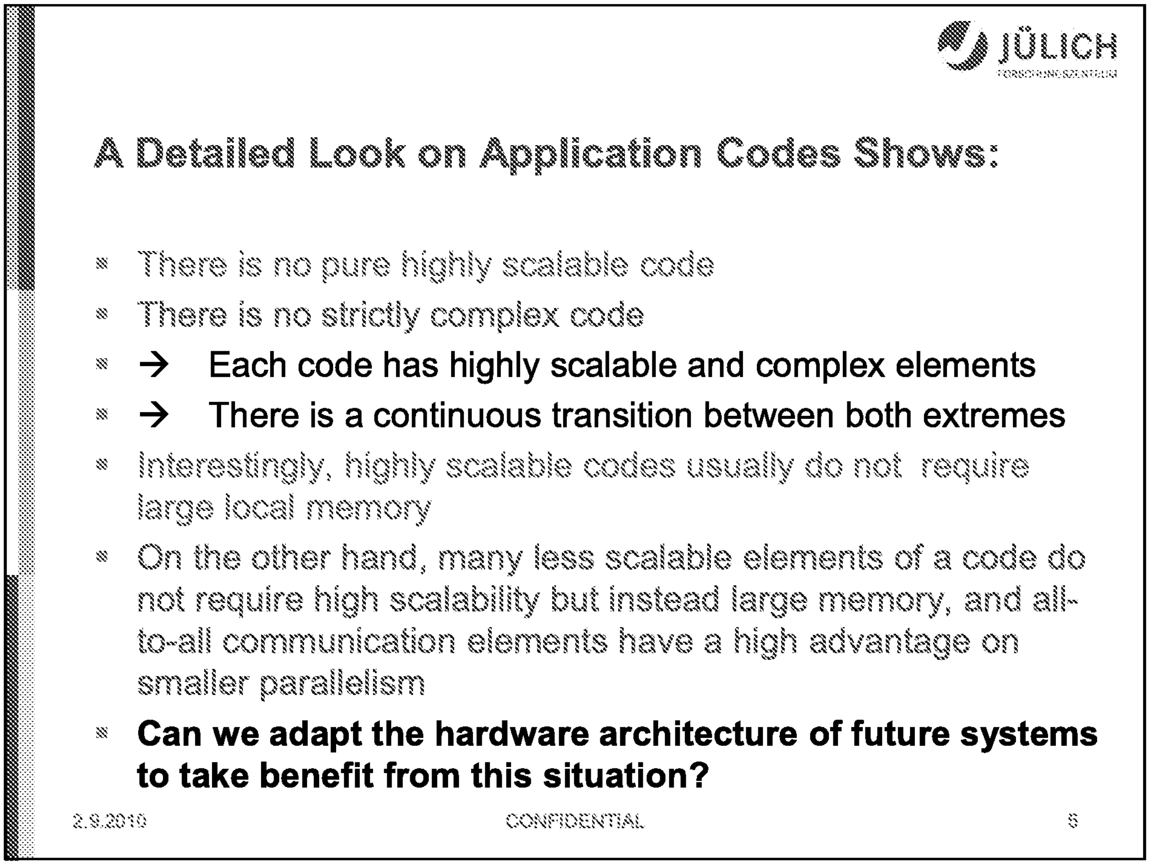
Intel Nehalem Clusters
JUROPA
200 TFlop/s
HPC-FF
100 TFlop/s

File Server
GPFS, Lustre

IBM Blue Gene/P
JUGENE, 1 PFlop/s

2.9.2010 CONFIDENTIAL 4

Exhibit E



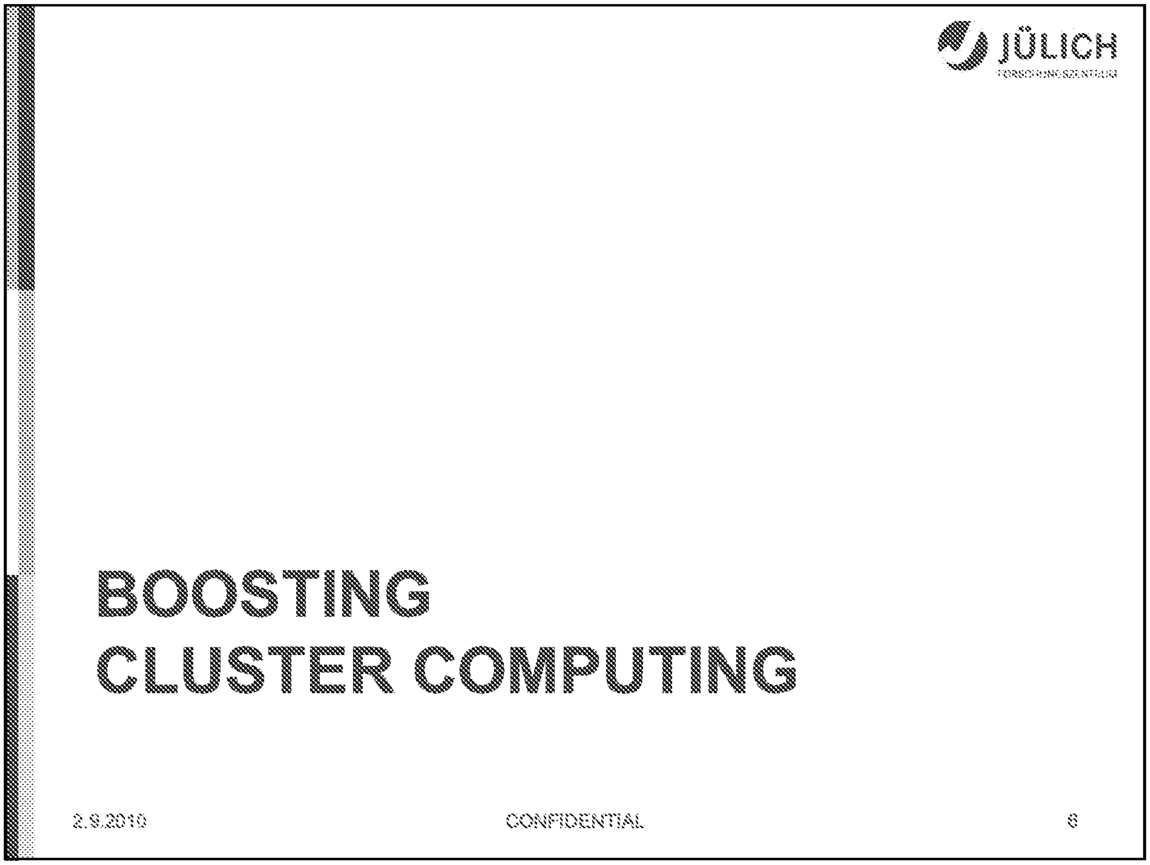
JÜLICH
FORSCHUNGSZENTRUM

A Detailed Look on Application Codes Shows:

- There is no pure highly scalable code
- There is no strictly complex code
- → Each code has highly scalable and complex elements
- → There is a continuous transition between both extremes
- Interestingly, highly scalable codes usually do not require large local memory
- On the other hand, many less scalable elements of a code do not require high scalability but instead large memory, and all-to-all communication elements have a high advantage on smaller parallelism
- **Can we adapt the hardware architecture of future systems to take benefit from this situation?**

2.9.2010 CONFIDENTIAL 5

Exhibit E

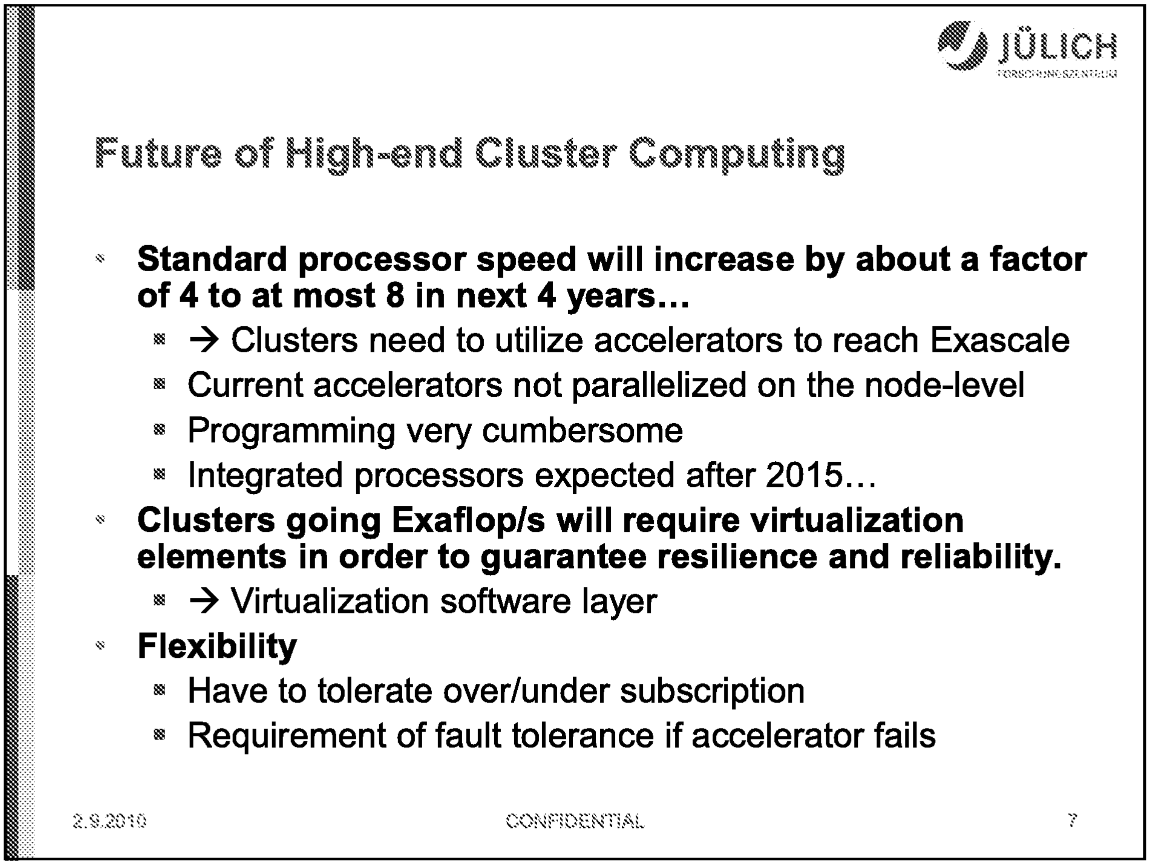


JÜLICH FORSCHUNGSZENTRUM

BOOSTING CLUSTER COMPUTING

2.9.2010 CONFIDENTIAL 6

Exhibit E



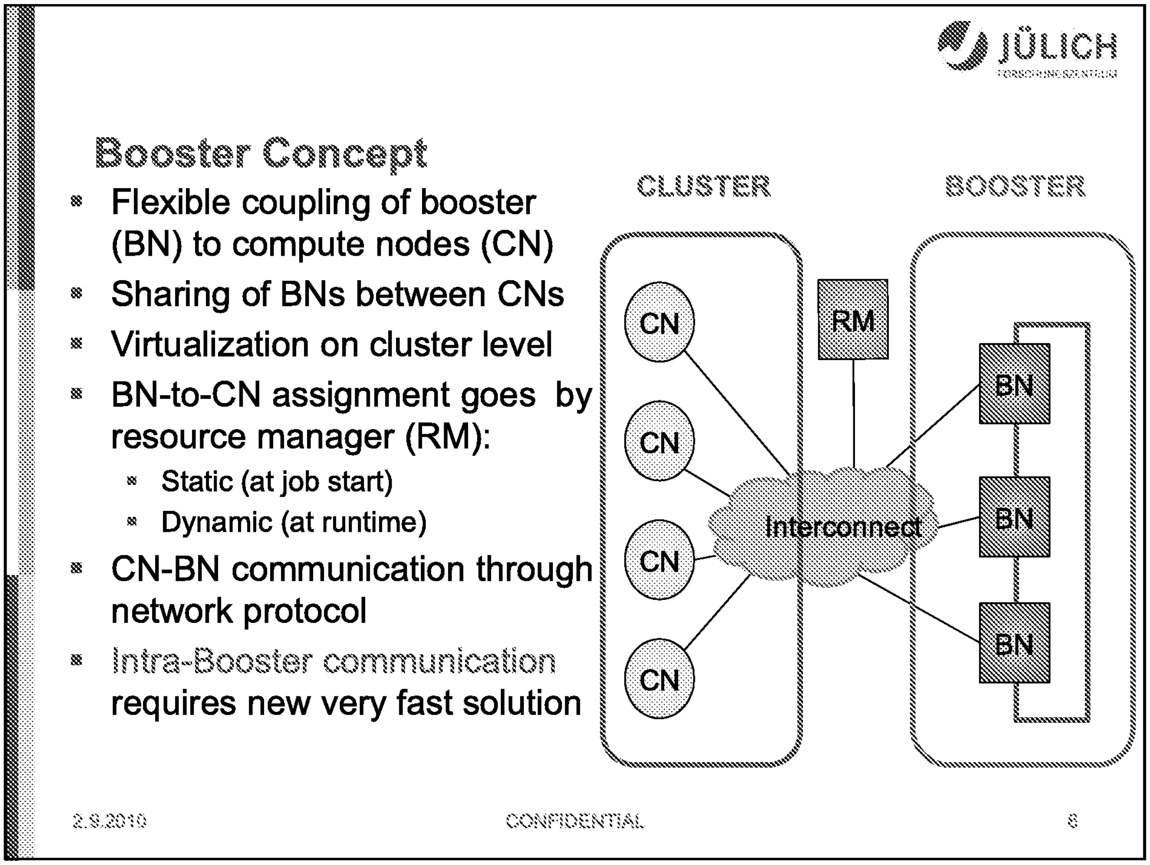
JÜLICH
FORSCHUNGSZENTRUM

Future of High-end Cluster Computing

- **Standard processor speed will increase by about a factor of 4 to at most 8 in next 4 years…**
  - → Clusters need to utilize accelerators to reach Exascale
  - Current accelerators not parallelized on the node-level
  - Programming very cumbersome
  - Integrated processors expected after 2015…
- **Clusters going Exaflop/s will require virtualization elements in order to guarantee resilience and reliability.**
  - → Virtualization software layer
- **Flexibility**
  - Have to tolerate over/under subscription
  - Requirement of fault tolerance if accelerator fails

2.9.2010 CONFIDENTIAL 7

Exhibit E



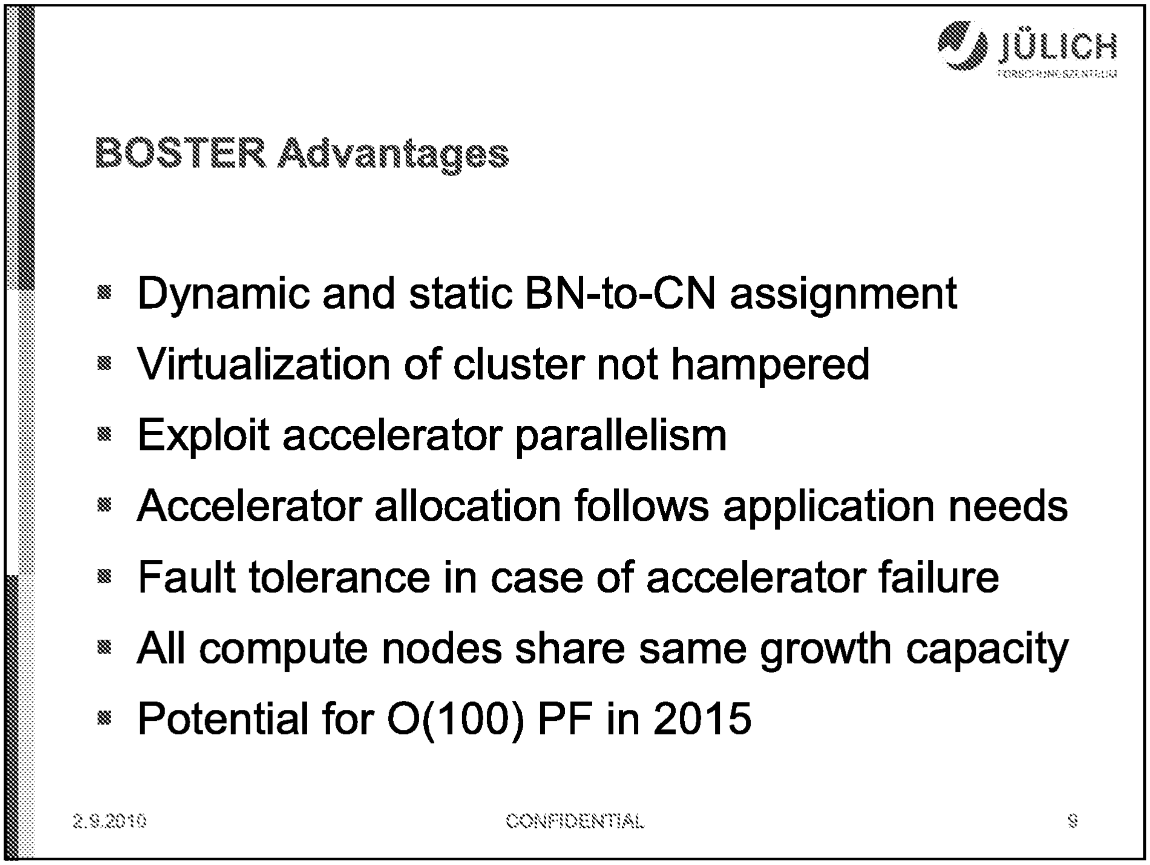
JÜLICH FORSCHUNGSZENTRUM

Booster Concept

- Flexible coupling of booster (BN) to compute nodes (CN)
- Sharing of BNs between CNs
- Virtualization on cluster level
- BN-to-CN assignment goes by resource manager (RM):
  - Static (at job start)
  - Dynamic (at runtime)
- CN-BN communication through network protocol
- Intra-Booster communication requires new very fast solution

2.9.2010 CONFIDENTIAL 8

Exhibit E



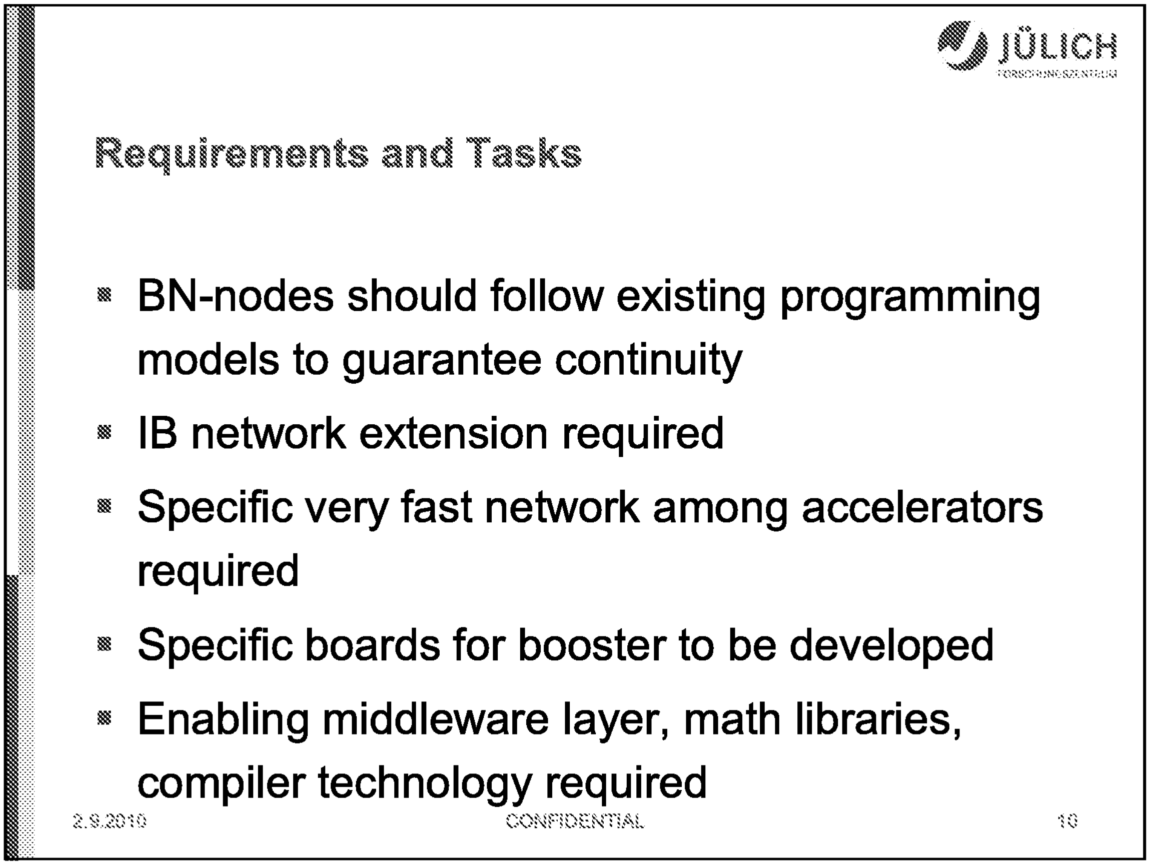
JÜLICH
FORSCHUNGSZENTRUM

BOSTER Advantages

- Dynamic and static BN-to-CN assignment
- Virtualization of cluster not hampered
- Exploit accelerator parallelism
- Accelerator allocation follows application needs
- Fault tolerance in case of accelerator failure
- All compute nodes share same growth capacity
- Potential for O(100) PF in 2015

2.9.2010 CONFIDENTIAL 9

Exhibit E



JÜLICH FORSCHUNGSZENTRUM

Requirements and Tasks

- BN-nodes should follow existing programming models to guarantee continuity
- IB network extension required
- Specific very fast network among accelerators required
- Specific boards for booster to be developed
- Enabling middleware layer, math libraries, compiler technology required

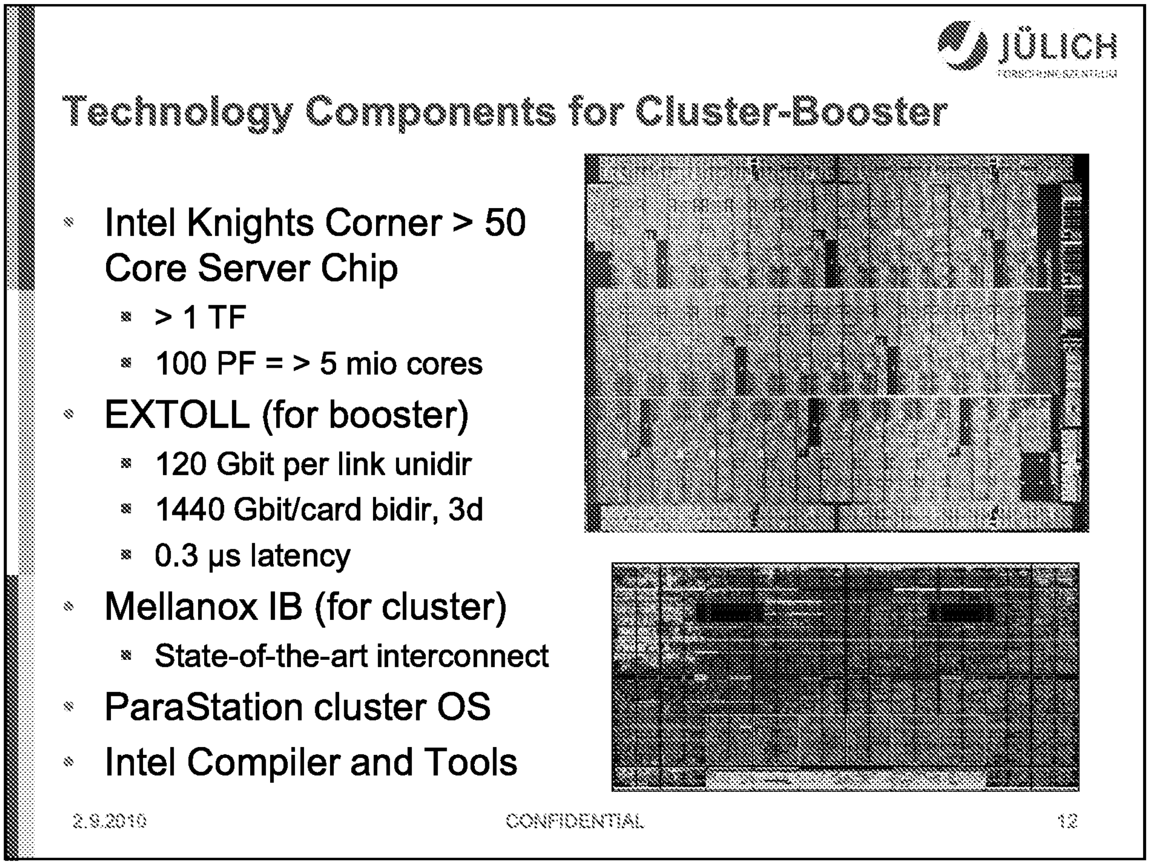
2.9.2010 CONFIDENTIAL 10

Exhibit E

PROPOSAL SKETCH

2.9.2010 CONFIDENTIAL 11

Exhibit E

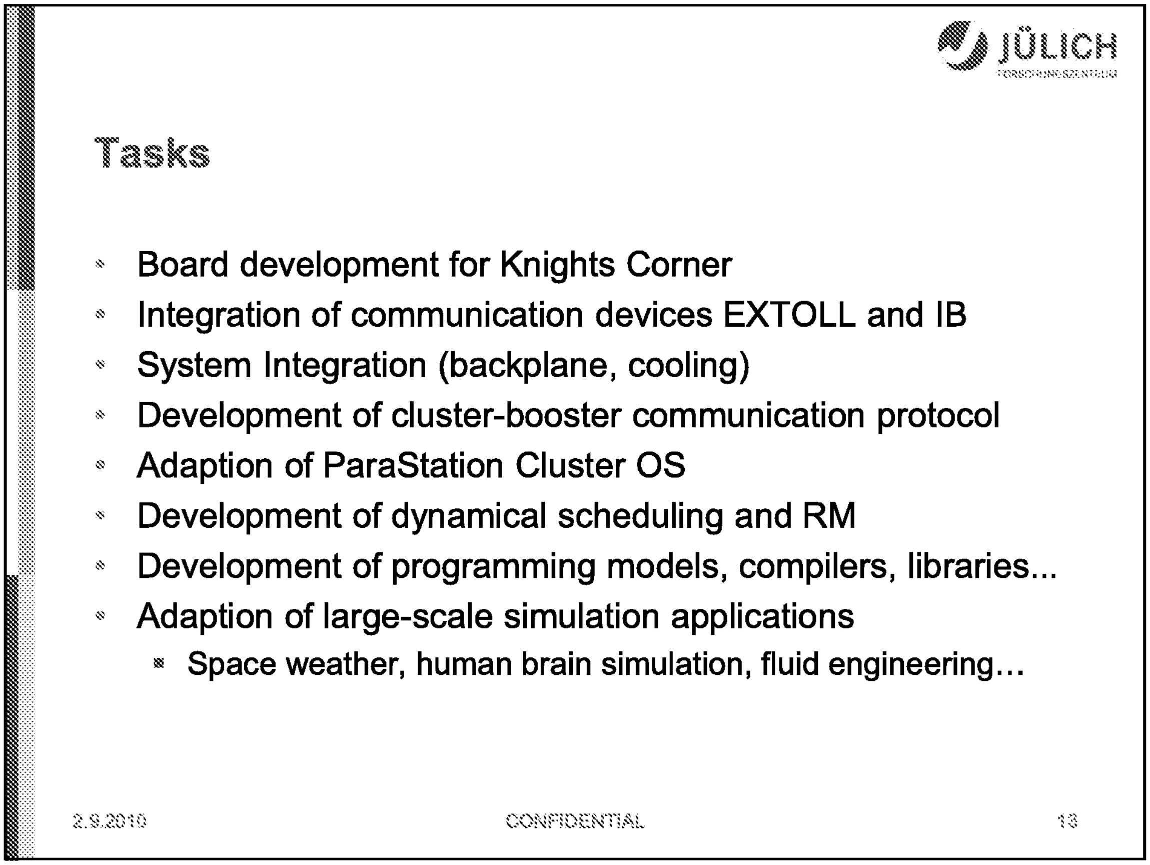


Technology Components for Cluster-Booster

- Intel Knights Corner > 50 Core Server Chip
  - > 1 TF
  - 100 PF = > 5 mio cores
- EXTOLL (for booster)
  - 120 Gbit per link unidir
  - 1440 Gbit/card bidir, 3d
  - 0.3 µs latency
- Mellanox IB (for cluster)
  - State-of-the-art interconnect
- ParaStation cluster OS
- Intel Compiler and Tools

2.9.2010 CONFIDENTIAL 12

Exhibit E

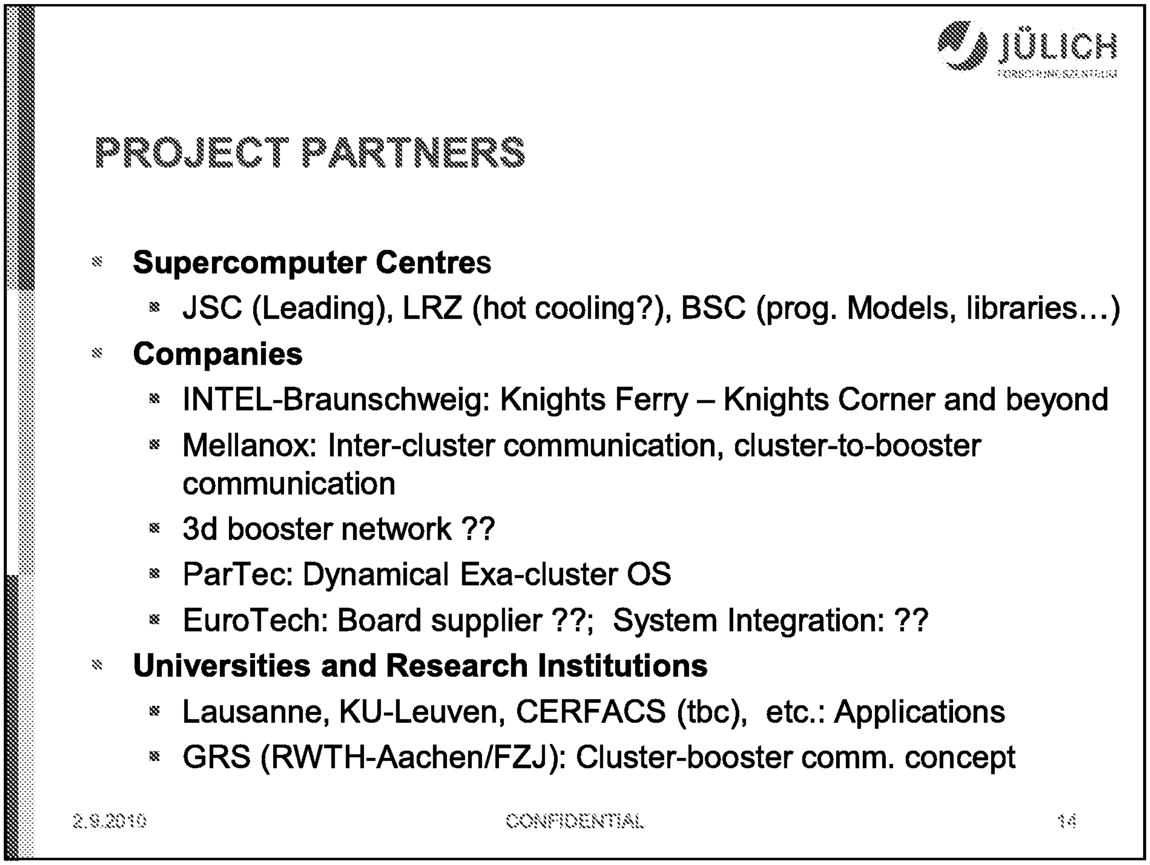


Tasks

- Board development for Knights Corner
- Integration of communication devices EXTOLL and IB
- System Integration (backplane, cooling)
- Development of cluster-booster communication protocol
- Adaption of ParaStation Cluster OS
- Development of dynamical scheduling and RM
- Development of programming models, compilers, libraries...
- Adaption of large-scale simulation applications
  - Space weather, human brain simulation, fluid engineering…

2.9.2010 CONFIDENTIAL 13

Exhibit E



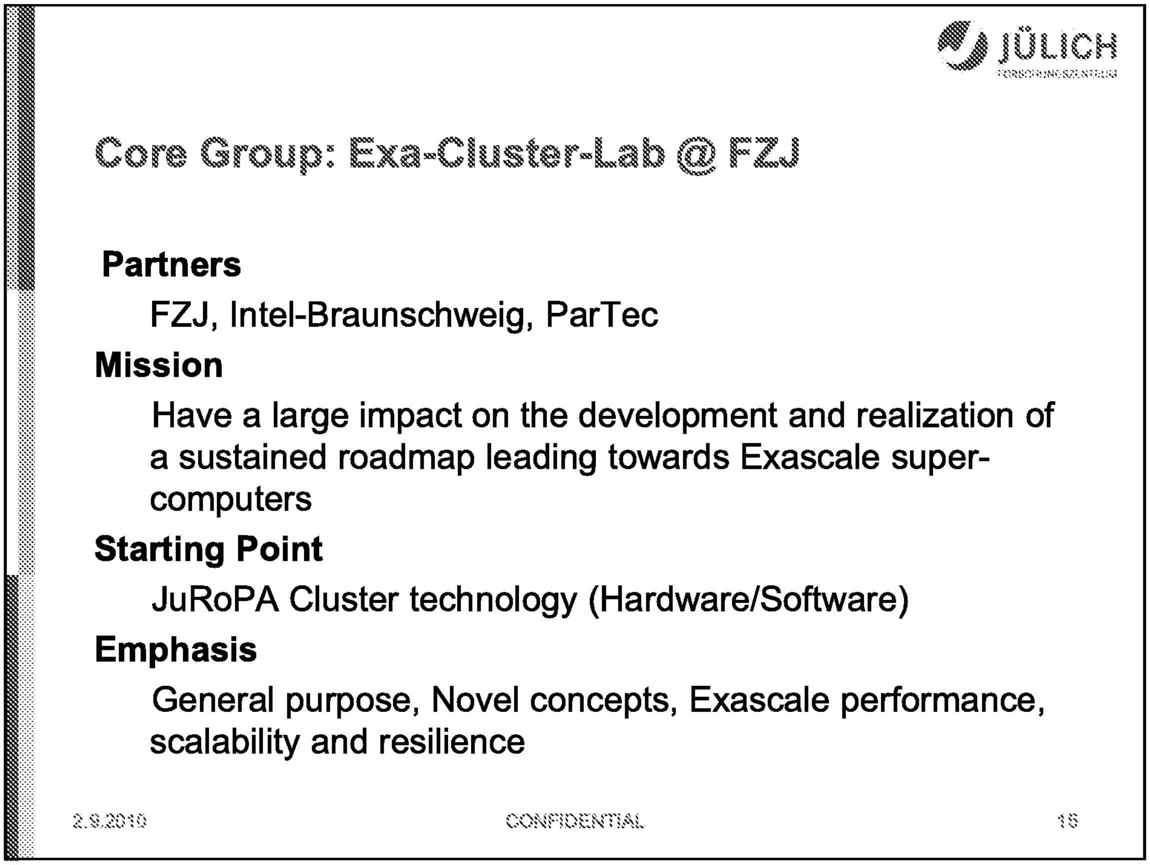
JÜLICH
FORSCHUNGSZENTRUM

PROJECT PARTNERS

- **Supercomputer Centres**
  - JSC (Leading), LRZ (hot cooling?), BSC (prog. Models, libraries…)
- **Companies**
  - INTEL-Braunschweig: Knights Ferry – Knights Corner and beyond
  - Mellanox: Inter-cluster communication, cluster-to-booster communication
  - 3d booster network ??
  - ParTec: Dynamical Exa-cluster OS
  - EuroTech: Board supplier ??; System Integration: ??
- **Universities and Research Institutions**
  - Lausanne, KU-Leuven, CERFACS (tbc), etc.: Applications
  - GRS (RWTH-Aachen/FZJ): Cluster-booster comm. concept

2.9.2010 CONFIDENTIAL 14

Exhibit E



JÜLICH
FORSCHUNGSZENTRUM

Core Group: Exa-Cluster-Lab @ FZJ

Partners

FZJ, Intel-Braunschweig, ParTec

Mission

Have a large impact on the development and realization of a sustained roadmap leading towards Exascale super-computers

Starting Point

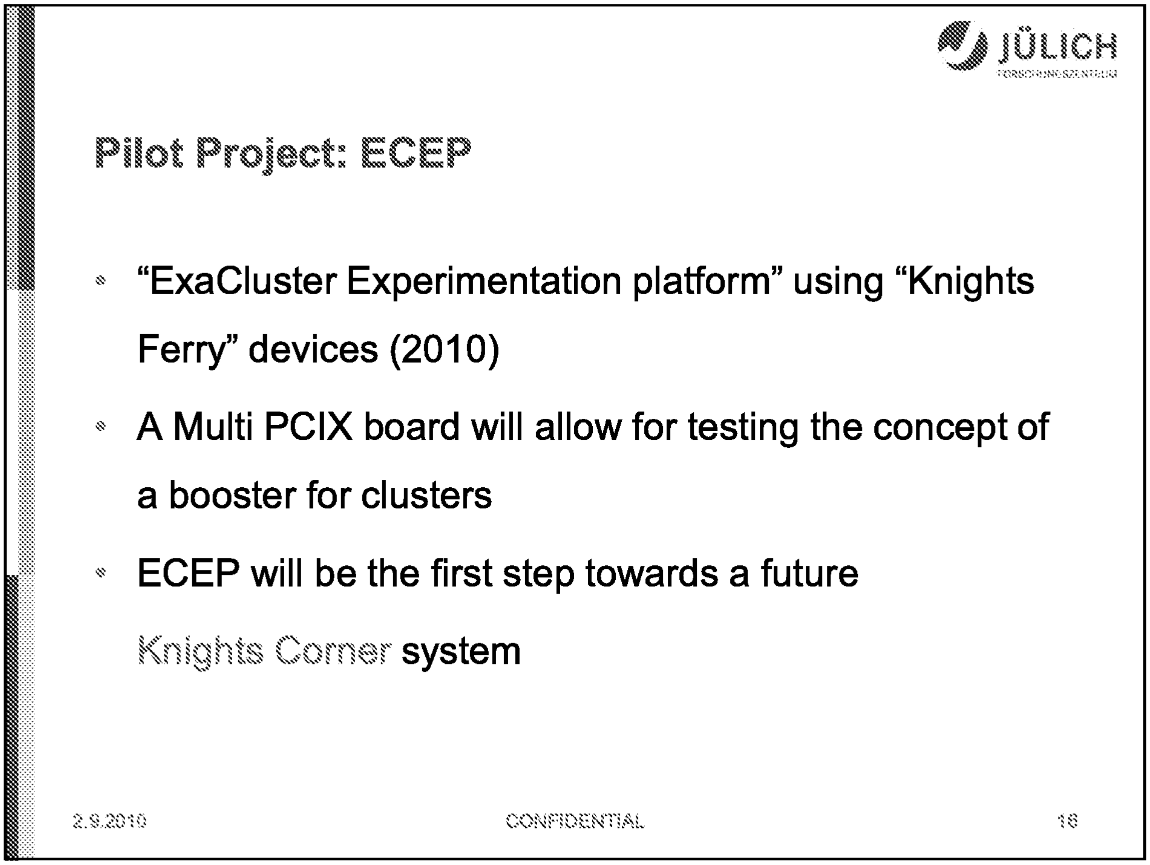
JuRoPA Cluster technology (Hardware/Software)

Emphasis

General purpose, Novel concepts, Exascale performance, scalability and resilience

2.9.2010 CONFIDENTIAL 15

Exhibit E

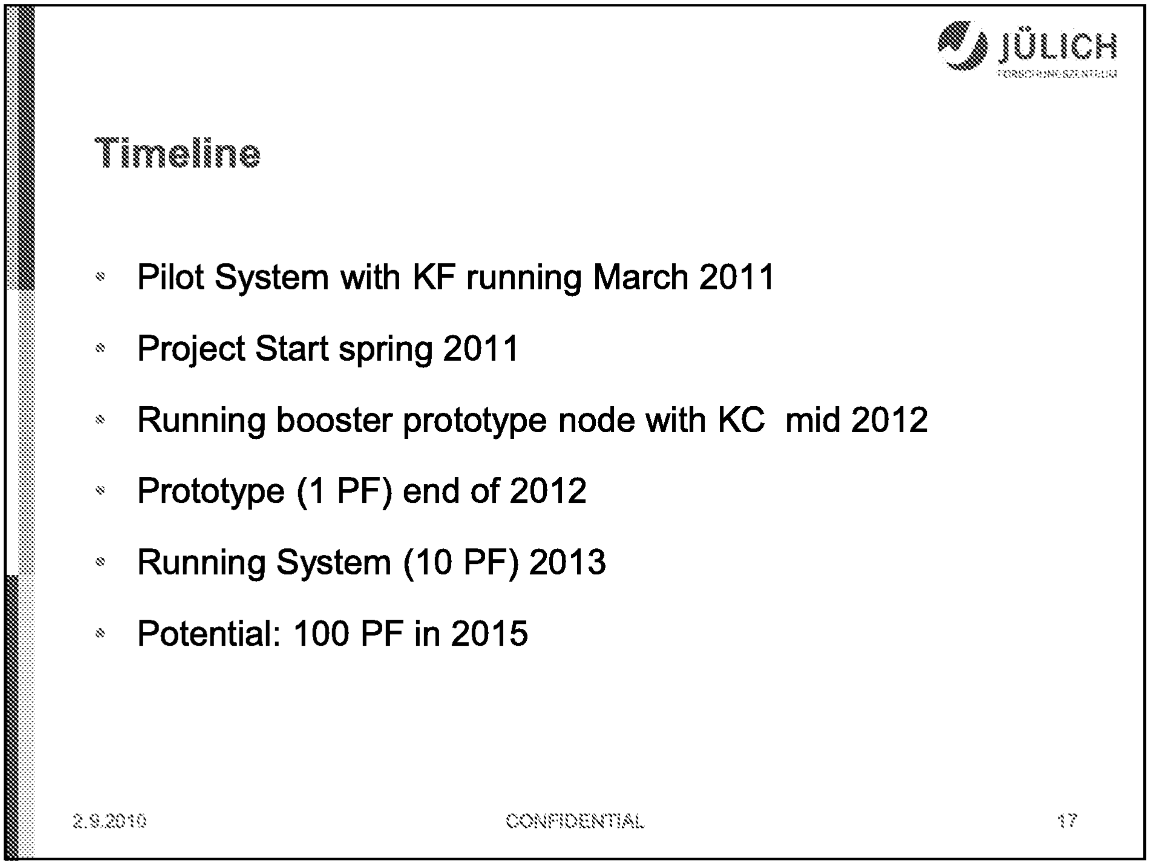


Pilot Project: ECEP

- “ExaCluster Experimentation platform” using “Knights Ferry” devices (2010)
- A Multi PCIX board will allow for testing the concept of a booster for clusters
- ECEP will be the first step towards a future Knights Corner system

2.9.2010 CONFIDENTIAL 16

Exhibit E



JÜLICH
FORSCHUNGSZENTRUM

Timeline

- Pilot System with KF running March 2011
- Project Start spring 2011
- Running booster prototype node with KC mid 2012
- Prototype (1 PF) end of 2012
- Running System (10 PF) 2013
- Potential: 100 PF in 2015

2.9.2010 CONFIDENTIAL 17